● **COMMENTARY**

Functional testing of ASD-associated genes

Marla B. Sokolowski[a,b,1]

There is a dire need for functional (causal) studies that move the field of neurodevelopmental disorders beyond statistical associations gleaned through genome-wide association studies (GWAS) and other "omic" approaches toward experimental manipulations of autism spectrum disorder (ASD)-associated genes and their genetic variants. ASDs are a genetically and phenotypically highly heterogeneous group of neurodevelopmental disorders with deficits in social interactions and communication, repetitive behaviors, and abnormalities in sensory processing (1). To date, >100 candidate genes and genetic variants are implicated in ASD through statistical associations, many of which are associated with other neurodevelopmental disorders. Functional tests of these ASD-associated genes can provide a means to uncover mechanisms involved in the underlying etiologies of ASD. One way forward is to test gene function in model systems. These include mammalian models such as nonhuman primates and mice and more simple high-throughput models such as zebrafish, *Drosophila melanogaster*, and *Caenorhabditis elegans* as well as human induced pluripotent stem cells and organoids along with postmortem brain samples. These models each have their advantages and disadvantages; however, the integration of findings from these models will provide insight into the mechanisms underlying ASD. In PNAS, McDiarmid et al. (2) report an elegant scalable, high-throughput pipeline to test functions of ASD-associated genes in the simple model organism, *C. elegans*.

Testing the Functions of Human ASD-Associated Candidate Genes in *C. elegans*

C. elegans is well known as an inexpensive, unbiased, and highly efficient vector for ascertaining the function of human genetic variants (3). Greater than 50% of human genes have structurally and functionally conserved *C. elegans* orthologs. It has a fully sequenced and well-annotated genome, a complete connectome with 302 neurons, and extensive behavioral assays. *C. elegans* reproduce rapidly (3 d from egg to adulthood).

Mutant strains for the majority of *C. elegans* genes are available and on a single genetic background, and gene editing (CRISPR-Cas9) is accurate (4). Human genes can functionally replace *C. elegans* orthologs, and alterations in these genes (e.g., missense mutations) can be tested for function (3).

McDiarmid et al. (2) in PNAS systematically inactivated highly conserved *C. elegans* ASD-associate genes and quantified the phenotypic consequences using a machine vision system called the Multi-Worm Tracker (Fig. 1). Twenty-six phenotypes covering morphology, developmental time, movement patterns, sensitivity to a mechanosensory stimulus, and habituation learning were measured in >27,000 worms of 135 strains, each carrying a mutation in an established highly conserved *C. elegans* ortholog of an ASD-associated gene. Their pipeline revealed hundreds of unique and shared relationships between phenotype and genotype that fit with the heterogeneity of human ASD, ranging from severe developmental delays and uncoordinated movement to subtle deficits in sensory and learning behaviors. The phenotyping pipeline quantified numerous parameters involved in habituation, a simple form of learning that occurs when a neural circuit exhibits the plastic ability to decrease its response to repeated sensory stimuli. Habituation, measured as timing of the response to a mechanosensory stimulus in *C. elegans*, is impaired in a number of neurodevelopmental disorders including ASD (3). McDiarmid et al. (2) clustered the strains of *C. elegans* according to the degree of similarity of their phenomic profiles and discovered parallel gene networks. They validated the pathways using epistasis (gene–gene interaction) analyses, supporting the hypothesis that phenotypic heterogeneity maps to genetic variation. Parallel networks centered on several well-known human ASD-associated genes. Chromodomain helicase DNA-binding protein 8 (CHD8)•chromodomain helicase DNA-binding protein 7 (chd-7) and neuroligin 3 (NLGN3)•neuroligin 1 (nlg-1) were the basis of

---

aDepartment of Ecology & Evolutionary Biology, University of Toronto, Toronto, ON M5S 3B2, Canada; and bChild and Brain Development Program, Canadian Institute for Advanced Research, Toronto, ON M5G 1M1, Canada
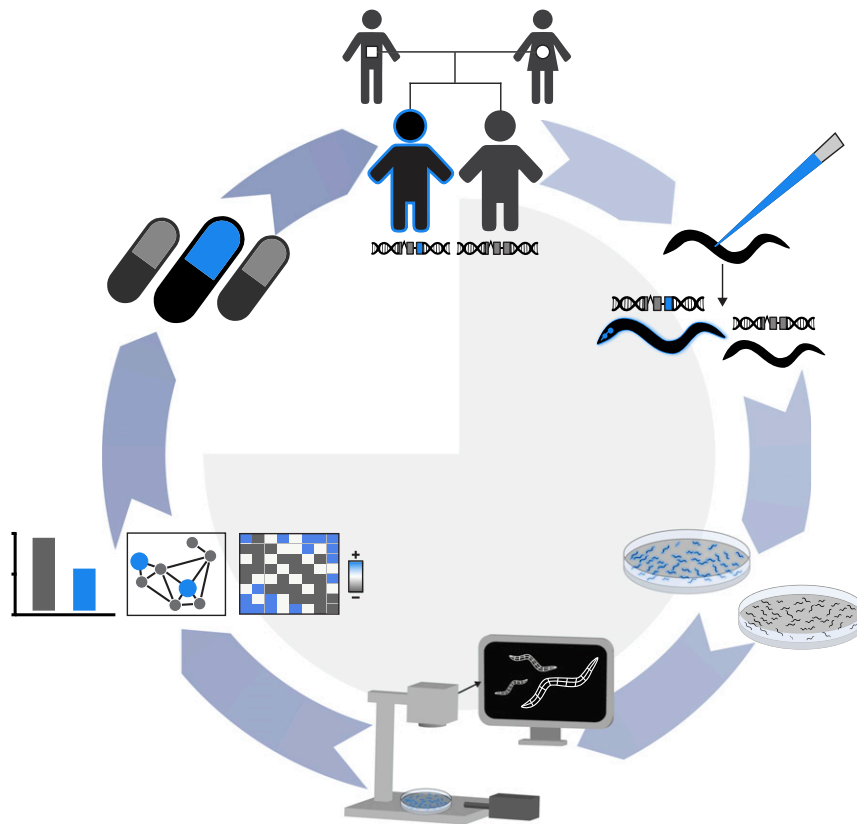[1]Email: marla.sokolowski@utoronto.ca.

Fig. 1. ASD-associated gene ortholog identification and phenomic characterization pipeline. *C. elegans* carrying mutations in orthologs of ASD-associated genes are engineered using CRISPR-Cas9 or obtained from stock centers. Large populations of *C. elegans* are grown, and phenotypic profiles are measured using the Multi-Worm Tracker. Strains are clustered based on phenotypic similarity identifying genotype–phenotype relationships, epistatic networks, and phenotypic reversibility. Reproduced with permission from ref. 3, which is licensed under CC BY 3.0.

mechanosensory hyperresponsivity and impaired habituation learning, respectively. Phenotypic profiles were also used to test the effects of missense variants. Finally, CRISPR-Cas9 auxin-inducible degradation demonstrated that the impaired habituation phenotype resulting from the developmental loss of NLG-1 could be partially rescued by expressing NLG-1 in the adult stage of development.

Lessons from Animal Model Behavior Genetics

It is instructive to consider some of the challenges facing ASD researchers through the lens of behavioral genetic research on model organisms (5, 6).

Behavioral Phenotypes Are Heterogeneous, Requiring Careful Characterization Prior to Genetic Analysis. Even in simple model organisms, complex behavioral phenotypes exhibit continuous variation that is not amenable to dichotomous categorization. Phenotypes associated with individuals on the ASD spectrum are also continuously distributed. However, most studies, whether based on genetics, neurobiology, or behavior, compare an ASD group to a typically developing group of individuals. Categorizing the behavioral heterogeneity of individuals within a population as typical or atypical does not take into account the heterogeneity of ASD. Despite the large overlap of continuously distributed data points between groups, group mean differences are still used to represent all individuals within a group. This approach does not lend itself to predictions of individual susceptibility or precision medicine. McDiarmid et al. (2)

successfully use a more dimensional fractionable approach to develop their phenomic data and embrace the heterogeneity of the phenotypes emerging from the behavioral phenotyping of worms carrying ASD-associated gene alterations. This approach recognizes the heterogeneity of the data and the variability within and between phenomic clusters.

There is Natural Genetic Variation for Most, If Not All, Individual Differences in Behavioral Traits in Animals. This statement is borne out through various methodologies, including pedigree analyses of families, artificial selection, quantitative genetics, candidate gene analysis, and, more recently, whole-genome sequencing technologies (GWAS). ASD has a strong genetic component, which includes monozygotic concordance estimates of ~70 to 90% and the identification of several distinct highly penetrant genetic syndromes (7).

Many Genes with Small Effects Influence Individual Differences in Behavior. Hundreds to thousands of genes can affect individual differences in behavior. One of the challenges is to identify which of these genes are core to the development and functioning of individual differences in a behavior (8). There are >100 ASD-associated genes, all of which have small effects on ASD, accounting for a very small percentage of the total number of ASD cases (e.g., most frequently mutated genes account for <1% of the total cases) (7). These rare de novo genetic variants associate with neuronal signaling and development, chromatin regulation, and synaptic function (9). In GWAS, a candidate single-nucleotide

polymorphism (SNP) that passes the significance threshold is commonly associated with the nearest gene along the DNA sequence; however, it cannot be assumed that a SNP affects the functioning of the nearest gene. Some SNPs are positioned within the coding region of a gene, resulting, for example, in a stop codon or missense mutation. In this case, if the SNP has a functional effect on the phenotype it will act through the gene that it has disrupted. However, it is not uncommon for genes to be embedded within other genes or for genes to be transcribed from the opposite DNA strand, so without functional validation it is not known whether or not the gene identified by the SNP found by association from the GWAS is in fact involved in the trait of interest. More challenging still is assigning a function to a SNP that does not reside within the coding region of a gene. A SNP may reside in a number of locations that are suggestive of regulatory functions such as the putative promoter region(s) or the 5′ or 3′ untranslated region of a gene. SNPs can also be in introns embedded within or outside of a gene or act over very long distances to affect a gene (e.g., in trans-regulatory elements). The extent to which genetic variants residing outside the coding regions of genes (e.g., in regulatory regions) play an important role in ASD remains to be determined.

Most Genes That Influence Behavior Have Multiple Functions; They Are Pleiotropic. It is a challenging task to dissect the multiple functions of a gene. For example, the *foraging* gene in the fruit fly *Drosophila* produces at least 21 transcripts expressed through 4 transcription start sites whose expression is temporally and spatially regulated. The modular structure of genes can provide a way forward to interrogate behavioral pleiotropy experimentally (10). When GWAS associates with more than one genetic variant (e.g., SNP) in a gene, the extent of pleiotropy combined with the complexities of gene structure make it necessary to validate the effect of each individual genetic variant. SNPs that associate with and map within the same gene may not have the same effect on function. Results from McDiarmid et al. (2) support this notion.

Genetic Background Strongly Affects Behavior. Genetic background is defined as the genotypes of all genes in the genome that may interact and influence the focal gene under study (6). In animal research, it is prudent to control for variation in genetic background. McDiarmid et al. (2) tested the functions of *C.*

elegans ASD-associated mutant genes in a common genetic background. Notably, even with extensive control of genetic background and the rearing and testing environment, they did not find evidence for separate phenotypic classes of the ASD-associated genes. In contrast, at the molecular level of investigation they found functionally distinct groupings of ASD-associated genes. When genes were clustered according to phenomics certain genes were closer to each other than other genes, reflecting molecular interactions. In humans, genetic background most likely influences core ASD genes to affect individual differences in ASD; however, little is currently known about how this occurs.

Gene–Environment Interactions on Behavioral Phenotypes Are Pervasive. Genetic variation predisposes individuals to behave in a certain way, and the environment (which itself is multidimensional) can modify this predisposition (11, 12). In model organisms, the modification of gene expression through experience can readily be seen at the molecular level acting through gene regulatory networks that act within circuits (13). Large-scale studies of the genetic contributions to ASD rarely consider environmental factors (14). Rutter's (14) reflection on gene–environment interplay in ASD states, "Whether or not such non-genetic factors will turn out to be measured environmental influences or, rather, stochastic (chance) effects remains to be seen, but both possibilities must be considered." ASD runs in families (9). If a family has one child with ASD, the full sibling of that child has a 10-fold increase in the risk of ASD. Much of this arises from genetic contributions, but unknown shared environments may also play a role. Future human research should investigate the interaction between allelic variation and gene expression and how the same ASD genetic risk factor can result in different outcomes. It is important to adopt a developmental approach using longitudinal studies of families that begin prior to pregnancy and cross the lifespan where detailed data about the environment/experience, genetics, epigenetics, imaging of brain development, and state-of-the-art measures of behavior are collected throughout development and adulthood (15).

When paired with human studies, model organisms such as *C. elegans* are well-positioned to contribute further knowledge about the molecular basis of gene function and the mechanisms underlying ASD (2).

1 American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, Washington, DC, ed. 5, 2013).

2 T. A. McDiarmid et al. Systematic phenomics analysis of autism-associated genes reveals parallel networks underlying reversible impairments in habituation. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 656–667 (2020).

3 T. A. McDiarmid et al., CRISPR-Cas9 human gene replacement and phenomic characterization in *Caenorhabditis elegans* to understand the functional conservation of human genes and decipher variants of uncertain significance. *Dis. Model. Mech.* **11**, dmm036517 (2018).

4 V. Au et al., CRISPR/Cas9 methodology for the generation of knockout deletions in *Caenorhabditis elegans*. *G3 (Bethesda)* **9**, 135–144 (2019).

5 K. S. Kendler, R. J. Greenspan, The nature of genetic influences on behavior: Lessons from "simpler" organisms. *Am. J. Psychiatry* **163**, 1683–1694 (2006).

6 M. B. Sokolowski, Drosophila: Genetics meets behaviour. *Nat. Rev. Genet.* **2**, 879–890 (2001).

7 L. de la Torre-Ubieta, H. Won, J. L. Stein, D. H. Geschwind, Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* **22**, 345–361 (2016).

8 E. A. Boyle, Y. I. Li, J. K. Pritchard, An expanded view of complex traits: From polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).

9 P. Szatmari, Risk and resilience in autism spectrum disorder: A missed translational opportunity? *Dev. Med. Child Neurol.* **60**, 225–229 (2018).

10 I. Anreiter, M. B. Sokolowski, The *foraging* gene and its behavioral effects: Pleiotropy and plasticity. *Annu. Rev. Genet.*, 10.1146/annurev-genet-112618-043536 (2019).

11 M. B. Sokolowski, D. Wahlsten, "Gene-environment interaction and complex behaviour" in *Methods in Genomic Neuroscience*, H. R. Chin, S. O. Moldin, Eds. (CRC Press, Boca Raton, FL, 2001).

12 M. J. Aristizabal et al., Biological embedding of experience: A primer on epigenetics. *Proc. Natl. Acad. Sci. U.S.A.*, 10.1073/pnas.1820838116 (2019).

13 D. F. Clayton et al., The role of the genome in experience-dependent plasticity: Extending the analogy of the genomic action potential. *Proc. Natl. Acad. Sci. U.S.A.*, 10.1073/pnas.1820837116 (2019).

14 M. Rutter, Changing concepts and findings on autism. *J. Autism Dev. Disord.* **43**, 1749–1757 (2013).

15 A. Thapar, M. Cooper, M. Rutter, Neurodevelopmental disorders. *Lancet Psychiatry* **4**, 339–346 (2017).